UNDERSTANDING THE DATASET

This project is based on an understanding of the factors to keep employees at the Company and which prompt others to leave.  The data can be downloaded from the [Dataset](https://www.kaggle.com/vjchoudhary7/hr-analytics-case-study) :  We need to use only  (3 files - General\_data.csv, Employee\_Survey\_Data.csv, Manager\_Survey\_data.csv) for the current project.

Let us try to understand each field of the data (general\_data.csv)

Below are the values each column has. The column names are pretty self-explanatory.

1.       AGE Numerical Value

2.       ATTRITION Employee leaving the company (0=no, 1=yes)

3.       BUSINESS TRAVEL (1=No Travel, 2=Travel Frequently, 3=Travel Rarely)

4.       DEPARTMENT (1=HR, 2=R&D, 3=Sales)

5.       DISTANCE FROM HOME Numerical Value - THE DISTANCE FROM WORK TO HOME

6.       EDUCATION Numerical Value. (1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor')

7.       EDUCATION FIELD (1=HR, 2=LIFE SCIENCES, 3=MARKETING, 4=MEDICAL SCIENCES, 5=OTHERS, 6= TECHNICAL)

8.       EMPLOYEE COUNT Numerical Value

9.       EMPLOYEE ID Numerical Value

10.   GENDER (1=FEMALE, 2=MALE)

11.   JOB LEVEL Numerical Value

12.   JOB ROLE (1=HR REP, 2=HR, 3=LAB TECHNICIAN, 4=MANAGER, 5= MANAGING DIRECTOR, 6= RESEARCH DIRECTOR, 7= RESEARCH SCIENTIST, 8=SALES EXECUTIVE, 9= SALES REPRESENTATIVE)

13.   MARITAL STATUS (1=DIVORCED, 2=MARRIED, 3=SINGLE)

14.   MONTHLY INCOME Numerical Value - MONTHLY SALARY

15.   NUMCOMPANIES WORKED Numerical Value - NO. OF COMPANIES WORKED AT

16.   OVER 18 (1=YES, 2=NO)

17.   PERCENT SALARY HIKE Numerical Value - PERCENTAGE INCREASE IN SALARY

18.   STANDARD HOURS Numerical Value - STANDARD HOURS

19.   STOCK OPTIONS LEVEL Numerical Value - STOCK OPTIONS (Higher the number, the more stock option an employee has)

20.   TOTAL WORKING YEARS Numerical Value - TOTAL YEARS WORKED

21.   TRAINING TIMES LAST YEAR Numerical Value - HOURS SPENT TRAINING

22.   YEARS AT COMPANY Numerical Value - TOTAL NUMBER OF YEARS AT THE COMPANY

23.   YEARS SINCE LAST PROMOTION Numerical Value - LAST PROMOTION

24.   YEARS WITH CURRENT MANAGER Numerical Value - YEARS SPENT WITH CURRENT MANAGER

b.      Let us try to understand about each field of the data (employee\_survey\_data.csv)

              1. Employee ID

2. Environment Satisfaction (1 'Low' 2 'Medium' 3 'High' 4 'Very High')

             3. Job Satisfaction (1 'Low' 2 'Medium' 3 'High' 4 'Very High')

             4. Work Life Balance  (1 'Bad', 2 'Good', 3 'Better', 4 'Best')

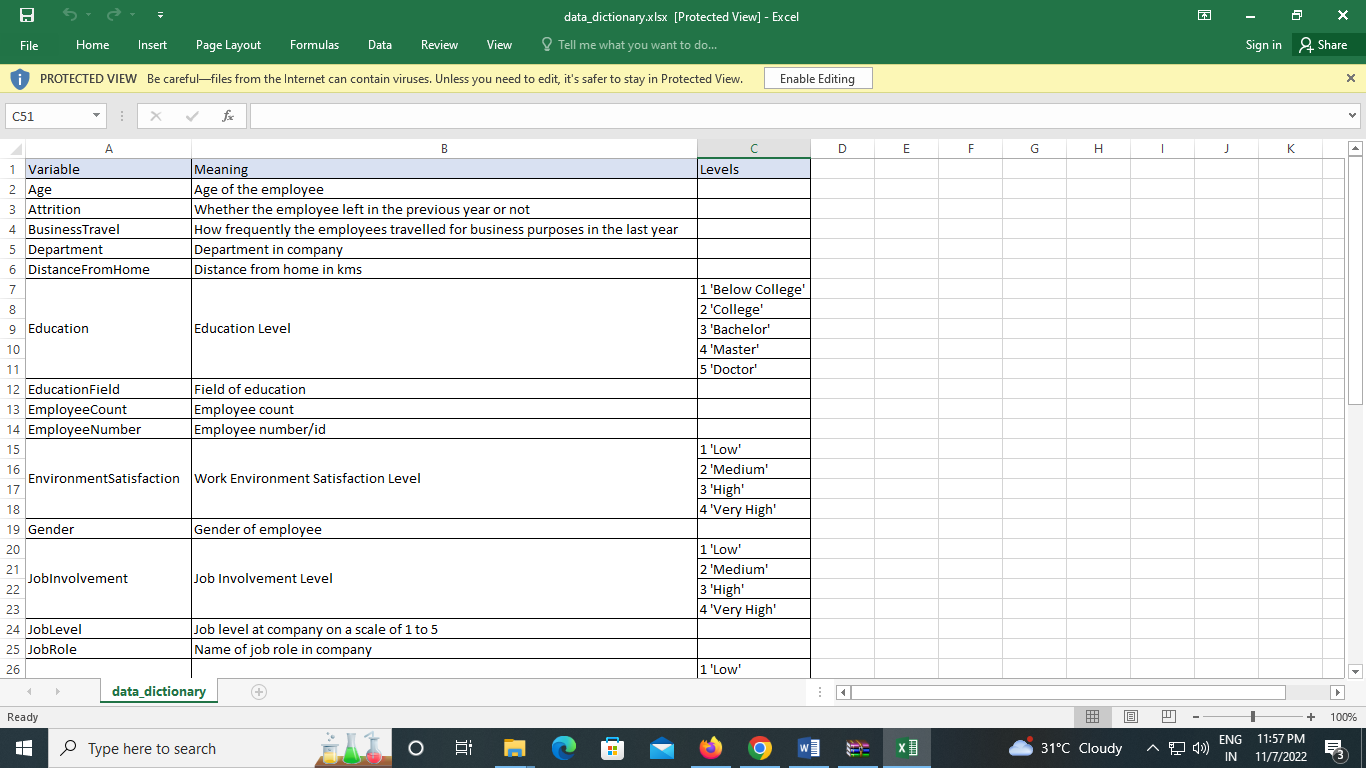
c.      Let us try to understand about each field of the data (manager\_survey\_data.csv)

         1. Employee ID

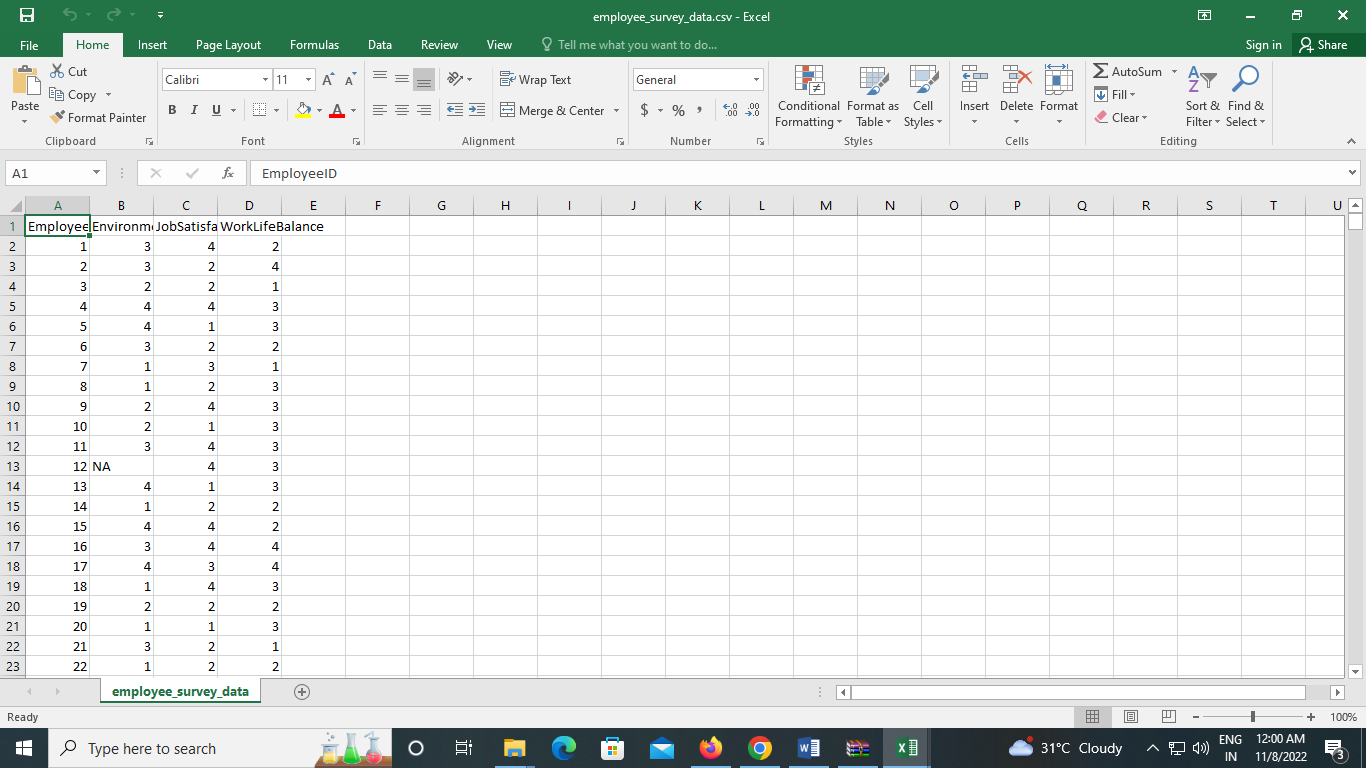
2. Job Involvement  (1 'Low' 2 'Medium' 3 'High' 4 'Very High')

3. Performance Rating ( 1 'Low', 2 'Good', 3 'Excellent', 4 'Outstanding')

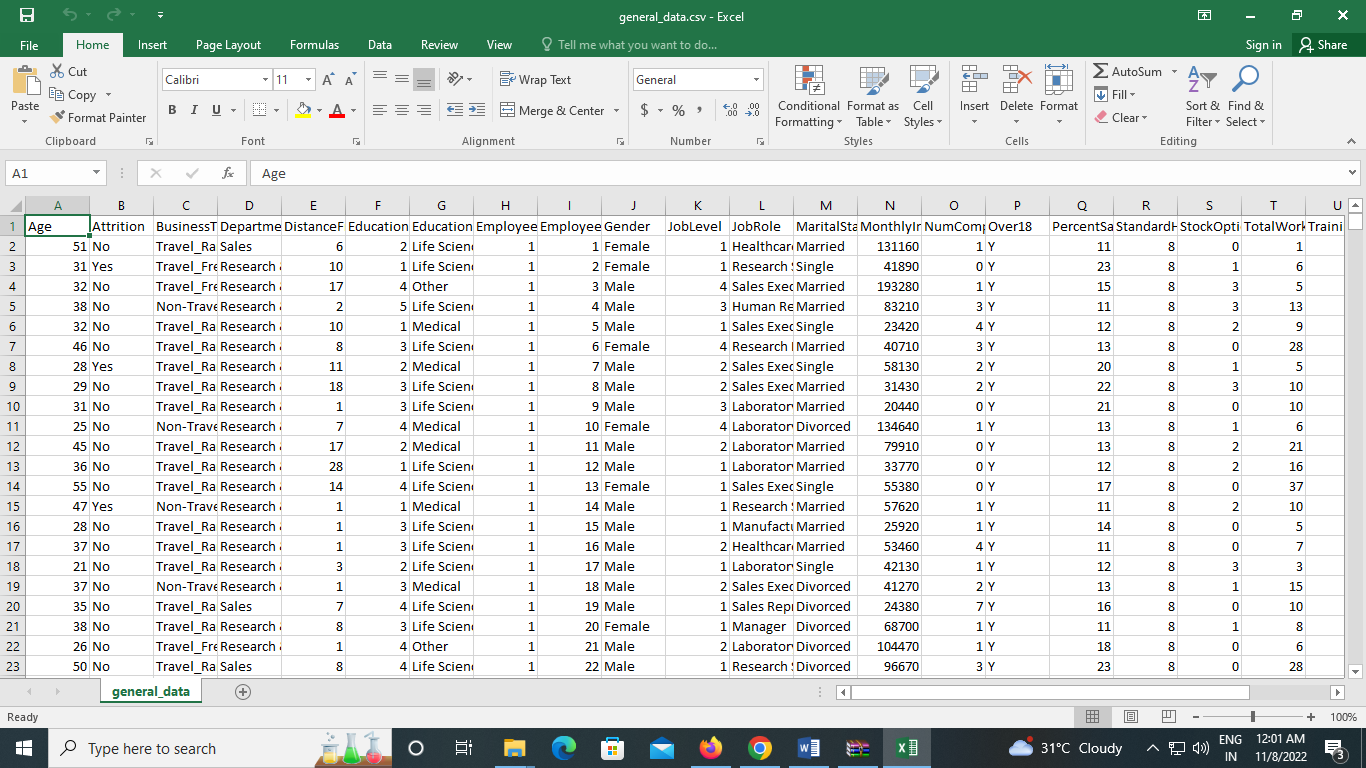
DATA DICTIONARY :



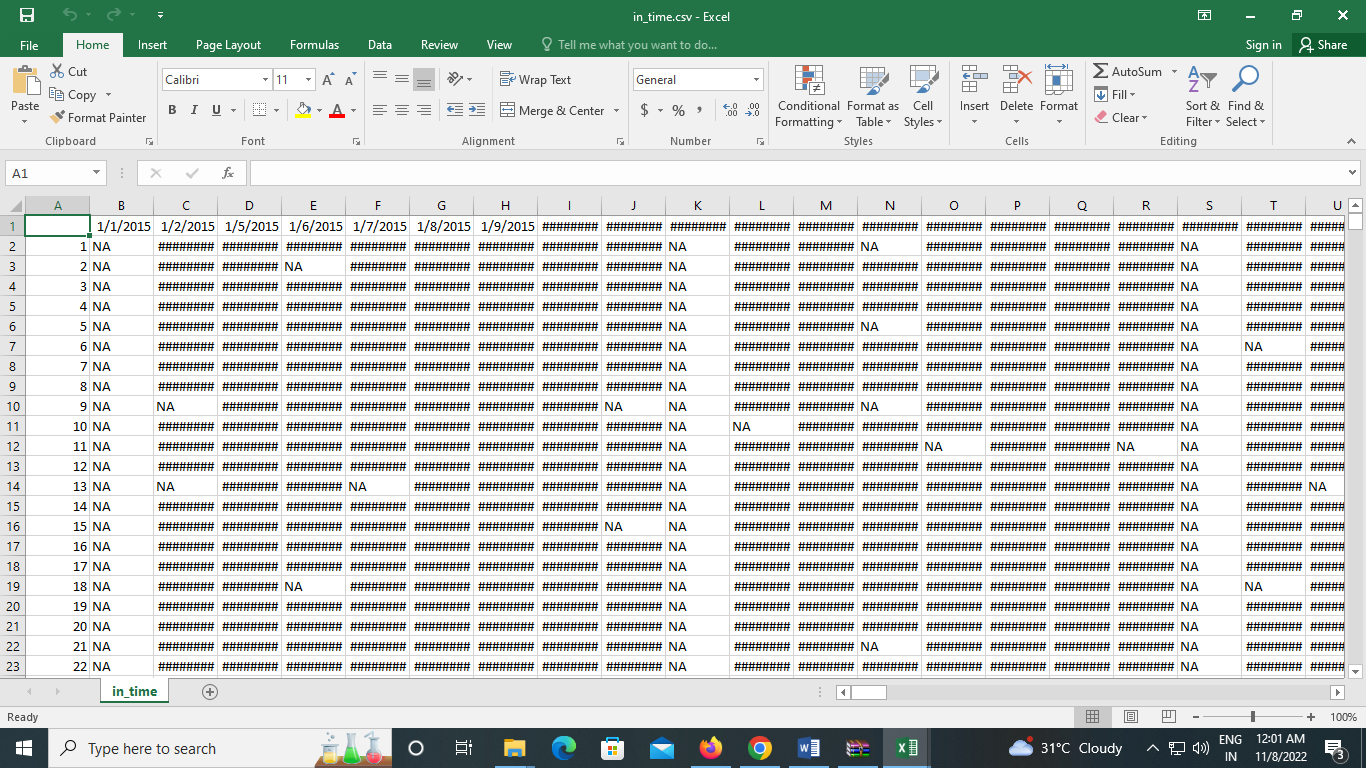
EMPLOYEE SURVEY DATA :



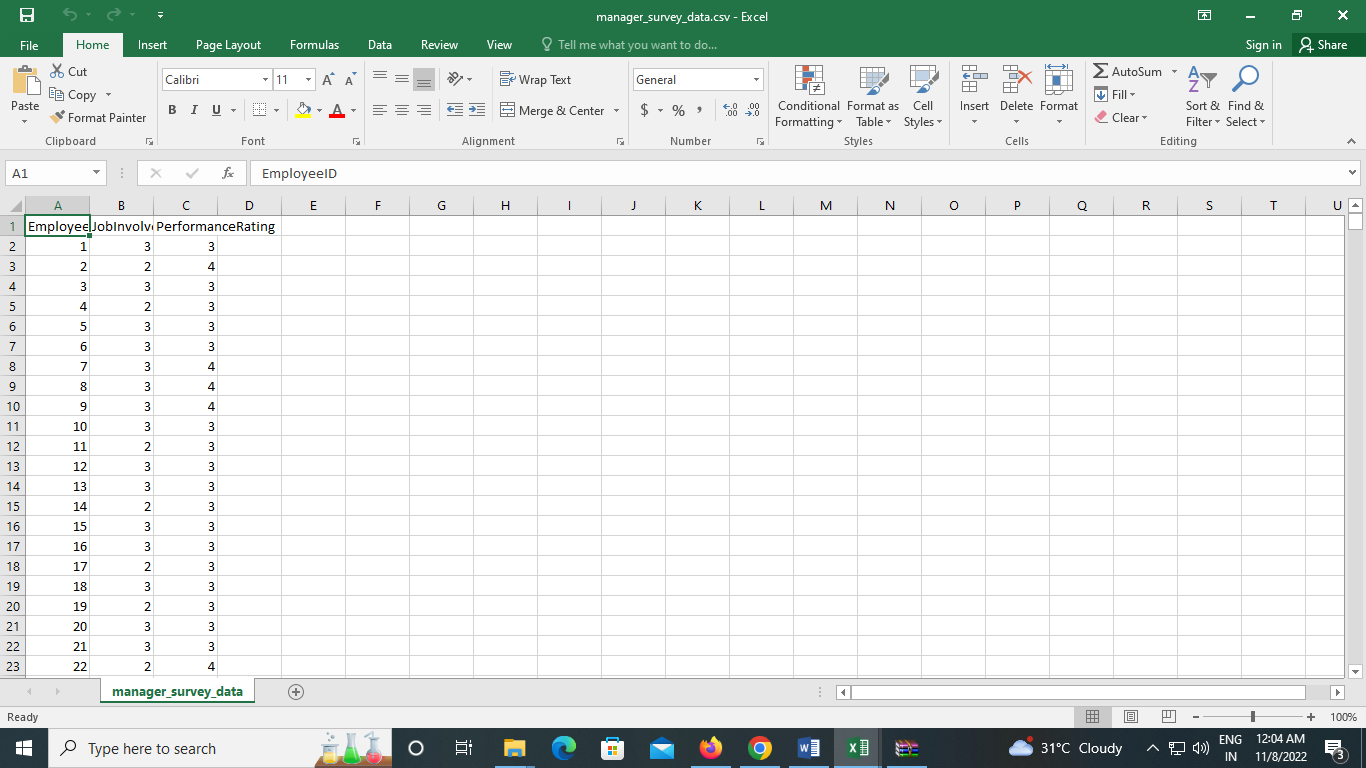
GENERAL DATA :



IN TIME :



MANAGER SURVEY DATA :



OUT TIME :

